

The development of vowel length as a subphonemic cue

Abigail Fergus, Kaitlyn Harrigan & Anya Hogoboom*

Abstract. Previous research has shown that English speakers use vowel length (VL) as a subphonemic cue to obstruent voicing. Many studies have demonstrated adults' ability to make a voicing judgment based upon VL but studies with children have provided mixed results. In the present study, we sought to first determine whether adults would exhibit varying sensitivity to VL based upon whether it could serve as a strong subphonemic cue. Second, we sought to better understand children's sensitivity to subphonemic VL from 4 to 6 years by removing top-down information and isolating the acoustic system. Adults (N=63) revealed greater sensitivity to VL preceding an obstruent showing that a subphonemic position boosts perceptibility for VL. Children from 4 to 6 (N=73, MEAN AGE=5;5.6) treat subphonemic VL differently from adults in two ways. First, they fail to show sensitivity at the same level as adults. 5- and 6-year-olds require VL differences that are twice as large and 4-year-olds do not show sensitivity even at the larger lengths. Second, children do not reveal varying sensitivity based upon the vowel's position as a potential subphonemic cue. This suggests that children have not fully developed their native phonology by the time they are 6-years-old.

Keywords. language acquisition; vowel length; subphonemic; voicing

1. Introduction. Historically, phonological theory has assumed that sounds belong to one of two distinct categories based upon their importance for meaning. The first of which being phonemic: acoustic properties that are critical for meaning (i.e. voicing) and the second being phonetic: properties that are not directly related to meaning (i.e. volume). But how does the perceptual system know whether a sound is voiced or voiceless? As it turns out, this is not an easy question to answer. Any phonemic distinction is cued by a number of acoustic cues (Repp 1982). For instance, voicing is cued by voice onset time (Andruski et al. 1994, McMurray et al. 2002, Pisoni & Tash 1974), vowel length (VL) (Denes 1955, Hogan & Rozsypal 1980, Miller & Dexter 1988, Port & Dalby 1982, Toscano & McMurray 2012, Warren & Marslen-Wilson 1988), F1 onset frequency (Stevens & Klatt 1974), and pitch (Haggard et al. 1970). These *subphonemic cues* are not themselves phonemic, but they cue the listener in to phonemic contrasts.

1.1. THE VOWEL LENGTH EFFECT. It is well documented in English that vowels preceding voiced consonants are longer in duration than those preceding voiceless consonants (Allen & Miller 1999, Chen 1970, House 1961, House & Fairbanks 1953, Kessinger & Blumstein 1998, Klatt 1973, Lisker 1957, Luce & Charles-Luce 1985, Sharf 1962). Following Ko et al. (2009), we will call this pattern the vowel length effect (VLE). Although it is not universal, the VLE has been documented in many of the world's languages (Chen 1970, Fintoft 1961, Fischer-Jgkgensen

* We are grateful to the members of the Child Language Lab and the Linguistics Research Group at William & Mary for input on earlier versions of this work. Authors: Abigail Fergus, William & Mary (abfergus@email.wm.edu), Kaitlyn Harrigan, William & Mary (kharrigan@wm.edu), and Anya Hogoboom, William & Mary (ahogoboom@wm.edu).

1964, Maddieson & Gandour 1975, Slis & Cohen 1969). Many perceptual studies have shown that English-speakers use VL as a cue to determine obstruent voicing (Denes 1955, Hogan & Rozsypal 1980, Port & Dalby 1982, Toscano & McMurray 2012, Warren & Marslen-Wilson 1988). The VLE seems to have primary importance determining the voicing of an obstruent in coda position (Klatt 1973, 1976, Sharf 1962, Denes 1955, Hogan & Rozsypal 1980, Warren & Marslen-Wilson 1988), word-medially (Port & Dalby 1982) and word-initially (Toscano & McMurray 2012). Thus, VL is a strong subphonemic cue to voicing in English.

1.2. DEVELOPMENT. Less is known about how children learn to use VL as a subphonemic cue. Research has generally indicated that young infants are able to show implicit discrimination of VL. However, their discrimination is less robust than adults and they are still beginning to develop awareness of the VLE. One such study by Eilers et al. (1984) found that 5- to 11-month-old English-learning infants were able to marginally discriminate duration differences, at a rate much lower than adults, and their discrimination improved as the duration difference became larger. A subsequent study tested discrimination by habituating infants to a non-word with a long (e.g. ta:ku) or short (e.g. taku) vowel. 18-month-olds were able to discriminate the non-words, thus showing sensitivity to VL in a subphonemic position (Mugitani et al. 2009).

Ko et al. (2009) tested whether infants expected VL and voicing to pattern in accordance with the VLE. They asked whether infants would detect a mismatch in VL and voicing by measuring their looking time. 8-month-olds did not show different rates of looking across matches and mismatches and 14-month-olds showed differences only when the vowel was long. However, a subsequent study tested infants' ability to recognize words with a subphonemic mismatch between voicing and VL (Swingley & Feest 2019). Infants did not fixate at different rates as a function of the mispronunciation showing that word recognition was not hindered by a subphonemic mismatch. Although we would not expect infants to reach a different lexical decision, it is nevertheless interesting that they do not show sensitivity to the mismatch.

Another line of work exploring children's development of subphonemic VL comes from their production. In general, production studies show that young children are relatively adept at producing the VLE. For instance, studies have revealed that children produce patterning consistent with the VLE at 21 months (Naeser 1970), 2 years (Ko 2007, Buder & Stoel-Gammon 2002), 3 years, and 4 years (Raphael et al. 1980). However, an earlier study by Stoel-Gammon & Buder (1999) reported that only 50% of the 2-year-olds produced patterns with the VLE suggesting that production of the VLE at younger ages might not be as robust.

Studies investigating children's perception of subphonemic VL have demonstrated mixed results. For instance, Greenlee (1980) tested children's voicing determinations of a word-final obstruent. Children listened to familiar words and indicated which word they heard of a minimal pair that differed only in the voicing of the final obstruent. For words with voiced final obstruents, voicing internal to the consonant was removed and the preceding vowel was shortened. For words with voiceless final obstruents the preceding vowel was lengthened. Thus, the researchers were able to determine if VL could cause participants to make voicing determinations that were contrary to the original voicing of the obstruent. They found that adults and 6-year-olds were able to make a voicing decision based on VL but 3-year-olds always responded according to original voicing.

However, a subsequent study by Wardrip-Fruin & Peach (1984) produced conflicting results. In their study, they also tested 3- and 6-year-olds and adults using English words ending in stops.

The stimuli were manipulated by removing various voicing cues to determine the relative weight of formant transitions, internal voicing, and preceding VL. Contrary to Greenlee (1980), they found that 3-year-olds weighed VL as a stronger cue to voicing than 6-year-olds and adults.

To add further complication, Krause (1982) found that 3- and 6-year-olds were able to make a voicing decision based upon VL when it was the only available cue to voicing but demonstrated a different threshold than adults. In order for children to switch from a voiceless to voiced decision, children required longer VLs than adults. The conflicting findings suggest that children's subphonemic VL may still be developing and is somewhat fragile at these ages.

1.3. **THE CURRENT STUDY.** This study seeks first to determine whether adults will be differentially sensitive to VL based upon its availability as a subphonemic cue. We asked not whether adults could make a voicing determination based upon VL but rather if the length would be more perceptible in a position where it is useful due to its status as subphonemic.

Second, we compare adult sensitivity to children, expanding the child literature in several ways. We chose to test sensitivity and not phonemic decision making to eliminate the top-down cues that would be involved in testing children with known words and mapping to meaning. This allowed us to isolate the acoustic system and gain a better understanding of children's subphonemic VL. The study itself consisted of a sound discrimination task compared across children and adults using non-words.

2. Adult study. The adult study was created to determine whether VL would be differentially perceptible based upon its availability as a subphonemic cue and serve as a comparison for the child study. We tested whether VL differences are more perceptible before an obstruent where adults have been shown to use it as a subphonemic cue to voicing. We hypothesize first that larger length differences will be more perceptible than smaller length differences overall and second that perceptibility will vary as a function of the length's usefulness as a cue to voicing.

2.1. **SUBJECTS.** Sixty-three ($N=63$) native English-speaking adults between the ages of 18 and 57 (MEAN: 21.83 years) were recruited from the William & Mary research participant pool (SONA) and social media. Those recruited through SONA received class credit for their participation.

2.2. **DESIGN.** The study consisted of a sound discrimination task where participants listened to pairs of phonemically identical non-words and then indicated their level of similarity. The design was 2×6 with **ENVIRONMENT** and **VL DIFFERENCE** manipulated as within-subjects factors.

ENVIRONMENT refers to the phoneme that followed the target-vowel (i.e. the vowel for which we manipulated its length). This allowed us to test the hypothesis that speakers will treat VL differently depending on its predictiveness. There were two environments: **OBSTRUENT** and **BASELINE** where the former was a position where VL is a strong cue and the latter is not. In the **OBSTRUENT** environment, the target-vowel was found in a location where English-speakers typically use it subphonemically: before an obstruent.

The difference in length of the target-vowels across the non-word pair was also manipulated as a within-subjects factor. The length of the target-vowel ranged from 80 to 180 ms using 20 ms increments. The **VL DIFFERENCE** was calculated as the absolute value of the target-vowel's length in the second word of the pair minus the target-vowel's length in the first word. There were six levels of **VL DIFFERENCE**: 0, 20, 40, 60, 80, and 100 ms. There were four pairs at each **VL DIFFERENCE** within each **ENVIRONMENT**. Both orders of every pair were included.

Participants were given a forced choice response based on the similarity of the words within the pair of *yes perfect match* and *no not quite*.

2.3. STIMULI. Test stimuli for the study consisted of 48 pairs of CVCV non-words created using MBROLA (Dutoit & Pagel 1995) and Praat (Boersma & Weenink 2018) software. Non-words were synthesized through MBROLA using the US1 voice and then concatenated into pairs. The first consonant in the word was 80 ms, the second consonant was 60 ms, and the final vowel was 170 ms to account for final lengthening. The final three phonemes within each condition were identical to provide a consistent environment for the target-vowel but the first consonant was varied to provide variety. The final three segments were [ikə] for OBSTRUENT and [inə] for BASELINE. The following initial consonants were randomly assigned to pairs [t, p, f, v, z, s, ʃ, tʃ, n, m, r, dʒ]. The two words across the pair were phonemically identical and the pitch of the second word was 20 Hz higher than the first in order to simulate two different “voices.” The interstimulus interval (ISI) between the words was 200 ms.

Practice stimuli were created using the same process but were instead CVC. There were eight practice pairs: four 0 ms difference and four 100 ms difference. The target vowel was [a] for four pairs and [i] for four pairs to provide variety. Segments for the initial consonant were [p, f, v, z, ʃ, tʃ, r, dʒ]. The final consonant was a nasal to provide a neutral environment.

2.4. PROCEDURE. The study was run remotely through Ibex farm (Drummond 2016), a web-based platform for deploying psycholinguistics experiments. Before the study began, participants were directed to a sound check to adjust their volume. There were two phases of the study: practice and test followed by a demographic form and debrief.

The practice phase consisted of eight non-word CVC pairs. After each pair was played, the participant received feedback. Participants could replay the practice words as many times as they liked before seeing the answer.

In the test phase, the 48 pairs of CVCV non-words were randomly ordered and participants were asked to respond whether the pronunciation of the second was the same as the first on a scale of *yes perfect match* and *no not quite*. Each test stimulus could only be played once.

2.5. RESULTS. We find that adults provide different rates of *yes perfect match* responses across the different length differences and environments. The proportion of *yes perfect match* responses in the adult study is listed in Table 1 below.

	VL DIFFERENCE (ms)					
	0	20	40	60	80	100
BASELINE	0.853	0.813	0.679	0.631	0.472	0.349
OBSTRUENT	0.845	0.833	0.687	0.544	0.329	0.202

Table 1. Proportion *yes perfect match* responses in adult study

We used generalized linear mixed effects models to analyze the results. These models are well suited for analyzing categorical data (Baayen 2007, Jaeger 2008). The reported models have random intercepts. These models predict the probability of a specific response (*yes perfect match*) in the different environments (see Agresti 2002, Jaeger 2008). We ran a mixed-effect logit model with *yes perfect match* responses as the dependent measure, VL DIFFERENCE (0, 20, 40, 60, 80, 100) and ENVIRONMENT (OBSTRUENT, BASELINE) as fixed effects, and SUBJECT as a random effect. The model indicated a main effect of VL DIFFERENCE [$X^2_{(5)} = 196.415, p < 0.0001$]

and no main effect of ENVIRONMENT [$X^2_{(1)} = 0.0648, p = 0.799$] We also found an interaction between VL DIFFERENCE and ENVIRONMENT [$X^2_{(5)} = 16.453, p < 0.001$].

In order to better understand what is driving the interaction, we ran pairwise comparisons, analyzing each level of VL DIFFERENCE separately. Pairwise comparisons revealed no significant differences between the environment at 0, 20, or 40 ms. However, participants are significantly less likely to respond *yes perfect match* in BASELINE compared to OBSTRUENT at 60 ($p = 0.0385$), 80 ($p = 0.0007$), and 100 ($p = 0.0001$) ms. Figure 2 below illustrates the proportion of *yes perfect match* responses.

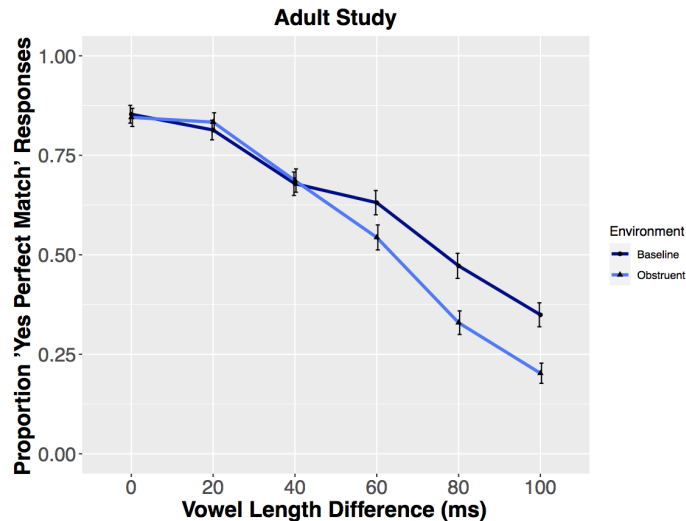


Figure 1. Proportion *yes perfect match* responses in adult study

2.6. DISCUSSION. The main effect of VL DIFFERENCE indicates that as VL differences became larger, participants were less likely to respond *yes perfect match* showing that they were more sensitive to the differences at these levels. This supports our hypothesis that participants will show greater sensitivity to greater VL differences. The interaction between VL DIFFERENCE and ENVIRONMENT indicates that participants are responding differently due to the combination of length and environment. Pairwise comparisons allowed us to identify that the interaction was being driven by the larger length differences (60, 80, and 100 ms). There were no significant differences in responses across environments at lower lengths (0, 20, and 40 ms) indicating that this difference is too small to warrant a *no not quite* response. Notably, a subphonemic position did not boost sensitivity enough to show different responses at these levels.

The differences at the higher VL differences indicates that VL is most perceptible in OBSTRUENT. Participants are least likely to give a *yes perfect match* response in OBSTRUENT, indicating that their sensitivity is influenced by potential subphonemic use providing support for our second hypothesis that a subphonemic position boosts perceptibility for VL.

3. Child studies. Seventy-six children between the ages of 4;0 and 6;11 were recruited remotely from across the United States. Children were recruited through the William & Mary Child Language Lab Database, the Cornell Play & Learning Lab Virtual Child Database, social media, local pre-schools, and the Cognitive Development Society listserv. Two children were excluded due to excessive background noise and one was excluded due to parental interference leaving 73 children in the study ($N=73$, MEAN AGE=5;5.6).

Four experiments were conducted to provide a comparison to the finding with adults showing boosted VL perception before obstruents. Piloting indicated that the range of within subjects factors in the adult study would create an overwhelming setup for children, leading us to simplify the adult design in several ways. These changes included: manipulating environment as a between subjects factor, pairing down the VL difference to two levels, using consistent pitch across the pairs, and using the same stimuli at practice and test.

3.1. EXPERIMENT 1: OBSTRUENT. Experiment 1 was created to serve as a direct comparison to the adult study. Because the literature suggests that children develop the phonological systems of their language by one year (Werker 1995), we predict that children will act similarly to adults. Specifically, we expect children to be sensitive to VL differences in this experiment because the length is found in a predictive position (before an obstruent).

3.1.1. SUBJECTS. Twenty-three children between the ages of 4;0 and 6;11 (MEAN = 5;5.77) recruited remotely from across the United States participated in Experiment 1.

3.1.2. DESIGN. The study consisted of a sound discrimination task deployed over Zoom where participants listened to pairs of phonemically identical non-words and were asked to indicate the level of similarity across the pair. The non-words were spoken by two robot characters and children indicated whether Baby Robot said it the *same* or a *little different* from Robot verbally or by pointing to one of two buttons on the top of the screen. Baby Robot always spoke second as he was attempting to copy Robot. VL DIFFERENCE was manipulated as a within-subjects factor.

The difference in length of the target-vowels across the non-word pair was manipulated as a within-subjects factor. VL DIFFERENCE was paired down to two levels: SAME and DIFFERENT. SAME corresponds to 0 ms and DIFFERENT corresponds to 100ms. For SAME, vowels were either both 80 ms or 180 ms resulting in a 0 ms difference across the pair. For DIFFERENT, one vowel was 80 ms and the other was 180 ms resulting in a 100 ms difference. The order of the vowels was varied. There were 10 SAME and 10 DIFFERENT pairs. Within SAME, 5 pairs had target-vowels of length 80 ms and 5 pairs had target-vowels of length 180 ms. Within DIFFERENT, both orders of every pair were included.

3.1.3. STIMULI. The test stimuli for the study consisted of 20 pairs of non-words taken from the obstruent environment of the adult study. The pairs were adjusted to the appropriate lengths (when necessary) in Praat. Unlike adults, practice stimuli were chosen from the test stimuli. There were six different practice pairs: two SAME and four DIFFERENT.

3.1.4. MATERIALS AND PROCEDURE. Children were tested remotely through Zoom. The study was a sound discrimination task constructed as a game where the children were helping Baby Robot practice his robot sounds. The experimenter first introduced the children to the characters and then explained the layout of the game. Children were told that to help Baby Robot learn his sounds, Robot was going to say a sound and Baby Robot was going to try to copy it exactly. They were then introduced to two buttons on the top of the screen, one red and one green, corresponding to whether Baby Robot said it the *same* or a *little different* from Robot.

The child then entered a practice phase. The first two practice trials were modeled by the experimenter and the child was instructed to listen carefully and the experimenter will tell them

whether Baby Robot said it the *same* or *different* by pressing the corresponding button¹. After the two modeled pairs, the child answered whether Baby Robot said it the *same* or *different* before the experimenter indicated the correct answer for the remaining four practice trials.

In the test phase, the child was “in charge” of which button was pressed. They received no feedback. After the child gave a response, the experimenter pressed the corresponding button. Responses were recorded by the experimenter on a score sheet.

3.1.5. RESULTS AND DISCUSSION. To analyze the results, we ran a mixed-effect logit model with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect. The model revealed no main effect of **VL DIFFERENCE** [$X^2_{(1)} = 1.878, p = 0.171$].

The lack of a main effect of **VL DIFFERENCE** shows that children were not sensitive to VL differences. Thus, the results did not support our hypothesis that children would be sensitive to VL differences before obstruents as was seen with adults. We identified two possibilities for why children were not sensitive to VL in this experiment. The first is that the context of the task is causing children to be less sensitive. In the adult experiment, participants were presented with pairs of non-words with no context. To make it age appropriate, we turned the task into a game with characters participating in a language exchange. Perhaps because the experiment closely resembled language, children were mapping to a lexical level while adults were not. Previous research suggests that linguistic context can influence sensitivity to phonemic contrasts (Werker & Tees 1984). This possibility was addressed in the second experiment where the language context was removed. The second possibility is that children have a different threshold than adults. This is addressed in the third experiment where the length differences are twice as large.

3.2. EXPERIMENT 2: OBSTRUENT, NON LANGUAGE. The purpose of the Experiment 2 was to test whether the context of the task might be driving differences between adults and children. We predict that children will perform equally in Experiment 2 and Experiment 1 indicating that a lexical-bias was not influencing their sensitivity.

3.2.1. METHOD. Eleven children between the ages of 4;0 and 6;11 (MEAN=5;5.7) recruited remotely from across the United States participated in Experiment 2.

The design was the same as the previous experiment except the pairs were played by a single computer instead of two robots. We used the same stimuli as Experiment 1 to isolate context and determine whether context was driving the lack of sensitivity.

Children were again tested remotely through Zoom. The experiment was a sound discrimination task constructed as a game where the children were helping Ollie the owl fix her computer. The children were told that Ollie created a computer that plays identical sounds. Lately, however, the computer hasn't been working very well and Ollie is calibrating it using two buttons. When the computer plays two sounds that are exactly the SAME, Ollie presses the green button. When the computer plays two sounds that are a little DIFFERENT, Ollie presses the red button. Once the child enters the test phase, they are asked to calibrate the computer for Ollie.

¹ This method of practice was implemented after the first 12 participants. Slightly different formats of practice were used previously but none were successful in helping children hear the VL differences.

3.2.2. RESULTS AND DISCUSSION. A mixed-effect logit model with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect showed no main effect of **VL DIFFERENCE** [$X^2_{(1)} = 0.489, p = 0.484$].

The lack of a main effect of **VL DIFFERENCE**, indicated that children were not sensitive to the VL differences in this experiment. Thus, removing the context did not improve children’s sensitivity allowing us to eliminate the possibility that the context of the task was causing children to be less sensitive.

3.3. EXPERIMENT 3: OBSTRUENT, LENGTHENED. Experiment 3 tested whether children have a different threshold than adults to show subphonemic use of VL. We predict that children will show sensitivity to VL with larger differences thus indicating a different threshold.

3.3.1. METHOD. Twenty children between the ages of 4;4 and 6;11 (MEAN=5;5.65) recruited remotely across the United States participated in Experiment 3.

The design was identical to Experiment 1. With the only difference being that the length difference was twice as large. For the SAME level, target-vowels were either both 80 ms or both 280 ms. At the DIFFERENT level, one target-vowel was 80 ms and the other was 280 ms. To create the stimuli, the VLs from the Experiment 1 stimuli were adjusted in Praat.

3.3.2. RESULTS. To analyze the data, We ran a mixed-effect logit model with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect. Unlike experiments 1 and 2, the model revealed a main effect of **VL DIFFERENCE** [$X^2_{(1)} = 34.708, p < 0.0001$], showing that participants are responding *same* at different rates across the SAME and DIFFERENT trials. Specifically, children are more likely to provide a *same* response when the VL is the SAME than when it is DIFFERENT. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 2 below.

SAME	DIFFERENT	<i>p</i> value
0.865	0.581	<0.0001

Table 2. Proportion *same* responses in Experiment 3

To determine if there were age-related differences, we ran three additional models by age (4, 5, and 6 years). All three models were mixed-effect logit models with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) as a fixed effect, and **SUBJECT** as a random effect. Interestingly, the models revealed a main effect of **VL DIFFERENCE** for the 6-year-olds [$X^2_{(1)} = 12.041, p < 0.0001$] and 5-year-olds [$X^2_{(1)} = 27.185, p < 0.0001$] but not for the 4-year-olds [$X^2_{(1)} = 0.000, p = 1.000$].

To determine why children failed to recognize VL differences in Experiment 1, we ran an additional mixed-effect logit model comparing all experiments in the OBSTRUENT environment (1, 2, and 3) with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) and **EXPERIMENT** (1, 2, 3) as fixed effects, and **SUBJECT** as a random effect. The model revealed a main effect of **VL DIFFERENCE** [$X^2_{(1)} = 35.425, p < 0.0001$]. The model did not find a main effect of **EXPERIMENT** [$X^2_{(1)} = 1.202, p = 0.548$]. Finally, the model revealed an interaction of **VL DIFFERENCE** and **EXPERIMENT** [$X^2_{(1)} = 16.779, p < 0.0001$].

We then ran pairwise comparisons analyzing each level of **VL DIFFERENCE** separately. At the SAME length, we find significant differences between Experiment 3 compared to Experiment 1 and 2 but no significant differences between Experiment 1 and Experiment 2. Compared to

Experiment 1, participants are significantly more likely to respond *same* on the SAME trials in Experiment 3 ($p = 0.044$) but not in Experiment 2 ($p = 0.410$). There were no significant differences between the experiments at the DIFFERENT level showing that the interaction was driven by differences in the SAME level. The proportion of *same* responses across the obstruent experiments (1, 2, and 3) at each VL difference is illustrated in Figure 2 below.

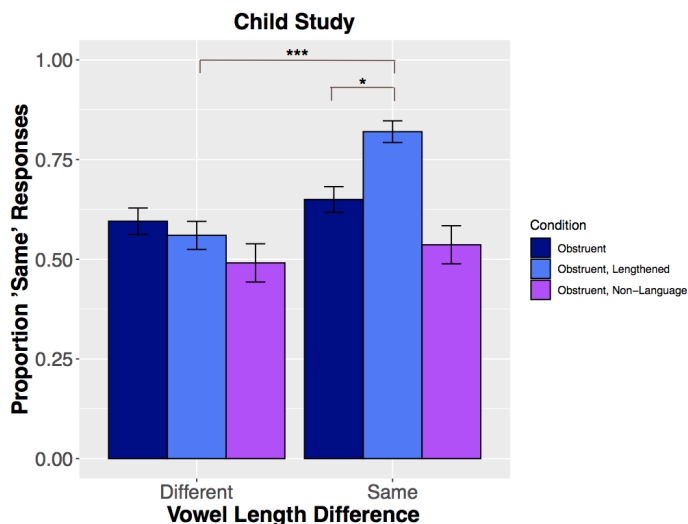


Figure 2. Proportion *same* responses in child experiments 1-3

3.3.3. DISCUSSION. The main effect of VL DIFFERENCE in Experiment 3 indicated that children require larger length differences to show sensitivity. The comparison between experiments 1, 2, and 3 indicated first that context was not driving the differences in responses and second that the larger length differences helped 5- and 6-year-old children succeed. However, even with the large length differences, 4-year-olds were not sensitive to VL. A comparison between Experiment 1 and 3 indicated that 100 ms is too little for 4- to 6-year-olds and 200 ms is enough for 5- to 6-year-olds to show sensitivity. It is unclear whether 4-year-olds might have an even larger threshold. However, 200 ms is already longer than what would be found in natural speech so if 4-year-olds did show sensitivity at a length larger than 200 ms, it would likely be due to a low-level acoustic strategy and not reflective of their developing phonology.

Interestingly, the different patterns of responses between Experiment 1 and 3 was driven by the SAME level not the DIFFERENT level. Suggesting that the increased threshold serves to make the SAME trials more salient. It is noteworthy, however, that this pattern is the opposite of what we found with adults. Recall, in the adult experiment the significant differences between environments were at the larger length levels (comparable to the DIFFERENT trials for children). This difference might suggest that 4- to 6-year-olds do not treat VL subphonemically.

To determine whether children are using VL subphonemically, they must show differing sensitivity based upon the potential usefulness, like adults. It is possible that their sensitivity is reflective of an acoustic skill and not their developing phonology. Experiment 4 investigated this.

3.4. EXPERIMENT 4: BASELINE, LENGTHENED. Experiment 4 was created to provide a comparison for Experiment 3 to determine whether children's sensitivity to VL is subphonemic. This experiment used the same lengthened VL differences from Experiment 3 but used the BASELINE environment instead of OBSTRUENT. If children perform equally well in this experiment, they likely do not use

VL subphonemically. However, if children perform worse in this experiment, they might be treating VL subphonemically at a larger threshold. We expect performance to be equal in this experiment because the larger threshold indicates that their sensitivity is likely phonetic.

3.4.1. METHOD. Twenty children between the ages of 4;5 and 6;9 (MEAN=5;5.2) were recruited remotely from across the United States to participate in Experiment 4.

The design was the same as Experiment 3 except the non-words were those from the BASELINE environment of the adult study instead of OBSTRUENT. To create the stimuli, BASELINE tokens from the adult study were manipulated in Praat to acquire the appropriate target-VLs. As with Experiment 3, the SAME level was 0 ms and the DIFFERENT level was 200 ms.

3.4.2. RESULTS. A mixed-effect logit model with *same* responses as the dependent measure, VL DIFFERENCE (SAME, DIFFERENT) as a fixed effect, and SUBJECT as a random effect revealed a main effect of VL DIFFERENCE [$X^2_{(1)} = 40.037, p < 0.0001$], showing that participants are responding *same* at significantly different rates across the SAME and DIFFERENT trials. Specifically, children are more likely to provide a *same* response when the VL is the SAME than when it is DIFFERENT. The proportion of *same* responses on the SAME and DIFFERENT trials is listed in Table 3 below.

SAME	DIFFERENT	<i>p</i> value
0.738	0.381	<0.0001

Table 3. Proportion *same* responses in Experiment 4

An additional mixed-effect logit model was run to compare experiments 3 and 4. This model used *same* responses as the dependent measure, VL DIFFERENCE (SAME, DIFFERENT) and EXPERIMENT (3, 4) as fixed effects, and SUBJECT as a random effect and revealed a main effect of VL DIFFERENCE [$X^2_{(1)} = 40.768, p < 0.0001$], and a significant (although much smaller) main effect of EXPERIMENT [$X^2_{(1)} = 3.863, p = 0.049$]. Finally, the model revealed no interaction of VL DIFFERENCE and EXPERIMENT [$X^2_{(1)} = 0.007, p = 0.978$] showing similar patterns of responses based upon the combination of EXPERIMENT and VL DIFFERENCE. The proportion of *same* responses in Experiment 3 and 4 at each length is shown below in Figure 4.

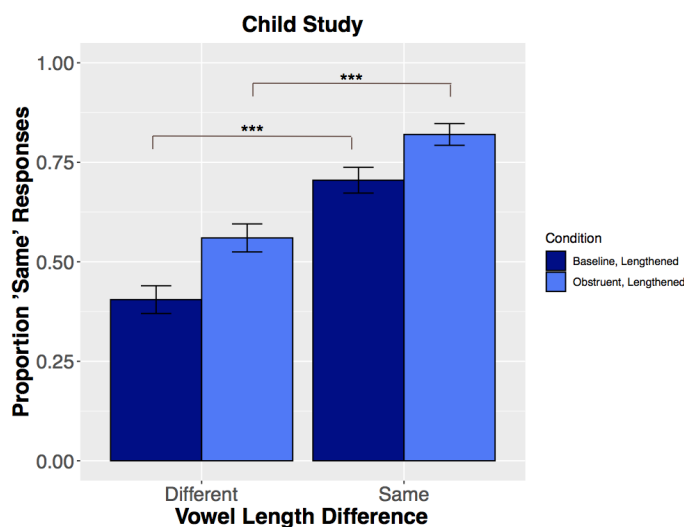


Figure 3. Proportion *same* responses in child experiments 3-4

Finally, to determine whether there were age-related differences, we ran three additional models divided by age. All three models were mixed-effect logit models with *same* responses as the dependent measure, **VL DIFFERENCE** (SAME, DIFFERENT) and **EXPERIMENT** (3, 4) as fixed effects, and **SUBJECT** as a random effect. The models revealed a main effect of **VL DIFFERENCE** for 6-year-olds [$X^2_{(1)} = 29.924, p < 0.0001$] and 5-year-olds [$X^2_{(1)} = 16.753, p < 0.0001$] but not for the 4-year-olds [$X^2_{(1)} = 2.754, p = 0.097$]. There was no main effect of **EXPERIMENT** and no interaction of **VL DIFFERENCE** and **EXPERIMENT** for any of the ages indicating similar patterns of responses based upon the combination of **EXPERIMENT** and **VL DIFFERENCE**.

3.4.3. **DISCUSSION.** The main effect of **VL DIFFERENCE** in Experiment 4 indicated that children show sensitivity to exaggerated VL differences in a neutral position. The main effect of **EXPERIMENT** in the model with both long experiments (3 and 4) showed that there were different rates of *same* responses between experiments 3 and 4. However, the lack of an interaction between **VL DIFFERENCE** and **EXPERIMENT** indicates that the different rates were not patterning based upon **VL DIFFERENCE**. In other words, children were not more or less likely to provide a *same* response based upon the VL difference in one experiment or the other. Because children were not treating VL differently in these two experiments, they did not show use of length as a subphonemic cue in the position preceding an obstruent. Instead, their sensitivity is likely phonetic since it is not influenced by its predictive potential. Thus, their sensitivity reflects a general acoustic strategy instead of their developing phonology.

Dividing the results by age revealed age-related differences in phonetic sensitivity but not subphonemic use. The lack of an interaction between **VL DIFFERENCE** and **EXPERIMENT** in all of the age groups indicates that none of the age groups were using VL subphonemically as they would need to be treating the length differently based on its position. Interestingly, however, the 4-year-olds did not show a main effect of **VL DIFFERENCE** indicating that they were not sensitive to VL differences in either experiment. Thus, 5- and 6-year-olds have a phonetic sensitivity to the larger VL differences in experiments 3 and 4 but 4-year-olds do not.

4. General discussion. Results from the adult and child study revealed that 4- to 6-year-old children treat VL quite differently than adults. Adults showed differential perceptibility of VL as a function of the VL's ability to inform the phonemic status of another sound. Children's sensitivity to VL differed from adults in three ways: by requiring a different threshold, being phonetic in nature not subphonemic, and showing differences in response rates at the **SAME** length level and not the **DIFFERENT** level. This pattern indicates that the VL served different purposes in the adult and child study.

5- and 6-year-old children's phonetic but not subphonemic sensitivity to VL in this study is consistent with work indicating that English-learning infants from 5 to 11 months (Eilers et al. 1984) and 18 months (Mugitani et al. 2009) are able to discriminate VL. Because the children required a larger threshold and were not treating VL differently as a function of its predictive potential, their ability to discriminate differences in experiments 3 and 4 was likely a result of low-level acoustic discrimination not their phonology. The explicit nature of this task builds off of previous findings by showing that 5- and 6-year-olds are able to show explicit discrimination of VL (as opposed to implicit in Eilers et al. 1984 and Mugitani et al. 2009). On the contrary, the study revealed that 4-year-olds are not able to demonstrate explicit discrimination of VL. However, their lack of explicit sensitivity does not necessarily mean that 4-year-olds lack the perceptual ability to discriminate VLs at this level. We expect 4-year-olds to have the capability

to hear these length differences since young infants are able to discriminate VL (Eilers et al. 1984, Mugitani et al. 2009) but they may have difficulty applying those skills in an explicit task.

We identify two potential explanations for why 4-year-olds did not show VL discrimination in this task. First, being cognitive overload. It is possible that the cognitive demand of the task is too great for 4-year-olds to discriminate VL differences. We suspect by testing 4-year-olds using an implicit measure and thus lowering the cognitive demand of the task will allow them to show sensitivity. Second, 4-year-old children might have a phonemic bias. Although Experiment 2 demonstrated that the context was not causing a lexical-bias, it's possible that hearing sounds displaying the phonotactics of English caused 4-year-olds to consider differences at the phonemic-level. It is quite adaptive for children at this age to pay particular attention to this level because as they are rapidly expanding their vocabularies, the phonemic level is the proper level at which to encode new vocabulary information. This explanation lies at the intersection of word recognition work showing that children's phonetic representations of familiar words are phonemic (Swingley & Aslin 2000) but not subphonemic (Swingley & Feest 2019) and word learning work showing that children's phonetic representations of new words is underspecified (Stager & Werker 1997).

The results found with children can also help to resolve some of the conflicting results we find when children are tested on their use of VL as a subphonemic cue to voicing (Greenlee 1980, Wardip-Fruin & Peach 1984, Krause 1982) and sensitivity to a subphonemic mismatch when recognizing familiar words (Swingley & Feest 2019). By using non-words that were not mapped to meaning, we were able to isolate the acoustic system and better understand how children treat VL in the absence of top-down information. It is difficult to reconcile why 3-year-olds but not 6-year-olds in Wardrip-Fruin & Peach (1984) and 6-year-olds but not 3-year-olds in Greenlee (1980) were able to make a voicing distinction based upon VL as they are in direct conflict with one another. However, our results suggest that even if children are weighing VL as a potential cue to voicing in these previous studies, their use of VL is not adult-like. Thus, children's subphonemic use of VL is not fully developed by 6 years. Finally, recall Krause (1982) showed that children required larger VL differences to make a voiced decision than adults. Our results are in some ways consistent with this finding but in some ways inconsistent. Krause (1982) showed that children needed larger VLs to reveal subphonemic use but in our study children needed larger VLs to show phonetic sensitivity. Together, both suggest that children need larger VL differences in general but the nature of their use differs across the studies. We suggest that the differing nature of children's use is a result of the presence (in Krause 1982) or lack (the present study) of top-down cues. Finally, the lack of subphonemic use found in our study is consistent with children's lack of subphonemic sensitivity in a word recognition task (Swingley & Feest 2019).

5. Future directions. This work lends itself to several directions for future inquiry. One being expanding to other languages. It would be of interest to compare speakers of languages that use the VLE as a cue to voicing and languages that do not to determine whether they show different patterns of sensitivity to VL. We would expect languages with the VLE to pattern similarly to English and languages without the VLE to not exhibit differences between BASELINE and OBSTRUENT. Further, we wonder whether speakers will show differing levels of boosted perceptibility in the OBSTRUENT environment dependent on the extent to which their language

exhibits and utilizes the VLE. Testing additional languages will allow us to explore other environments where VL is used subphonemically.

Another natural direction for this work is testing children using an implicit measure. Our results showing that children were unable to demonstrate explicit subphonemic sensitivity does not indicate that they are not sensitive at all. By testing children using an implicit measure such as eye-tracking, we would be able to better understand why children are failing to recognize subphonemic VL at the same threshold as adults in this study.

References

- Agresti, Alan. 2002. *Categorical data analysis*. New York: Wiley.
- Allen, J. Sean & Joanna Miller. 1999. Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *The Journal of the Acoustical Society of America* 106. 2031–2039. <https://doi.org/10.1121/1.427949>.
- Andruski, Jean, Sheila Blumstein & Martha Burton. 1994. The effect of subphonetic differences on lexical access. *Cognition* 52(3). 163–187. <https://doi.org/10.1121/1.427949>.
- Baayen, R. Harald. 2007. *Analyzing linguistic data: A practical introduction to statistics*. Cambridge: Cambridge University Press.
- Boersma, Paul & David Weenink. 2018. *Praat: Doing phonetics by computer* (Version 6.1.09). <http://www.praat.com/>.
- Buder, Eugene & Carol Stoel-Gammon. 2002. American and Swedish children's acquisition of vowel duration: Effects of vowel identity and final stop voicing. *The Journal of the Acoustical Society of America* 111(4). 1854–1864. <https://doi.org/10.1121/1.1463448>.
- Chen, Matthew. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22(3). 129–159. <https://doi.org/10.1159/000259312>.
- Denes, Peter. 1955. Effect of duration on the perception of voicing. *The Journal of the Acoustical Society of America* 27(4). 761–764. <https://doi.org/10.1121/1.1908020>.
- Drummond, Alex. 2016. *Ibexfarm* (Version 0.3.7). <https://github.com/addrummond/ibexfarm>.
- Dutoit, Thierry & Pagel, V. (1997). *MBROLA synthesizer* (Version 3.02b). <https://github.com/numediart/MBROLA>.
- Eilers, Rebecca, Dale Bull, D. Kimbrough Oller & Diana Lewis. 1984. The discrimination of vowel duration by infants. *The Journal of the Acoustical Society of America* 75(4). 1213–1218. <https://doi.org/10.1121/1.390773>.
- Fintoft, Knut. 1961. The duration of some Norwegian speech sounds. *Phonetica* 7(1). 19–39. <https://doi.org/10.1159/000258096>.
- Fischer-Jgkgensen, Eli. 1964. Sound duration and place of articulation. *STUF - Language Typology and Universals* 17(1–6). 175–208. <https://doi.org/10.1524/stuf.1964.17.16.175>.
- Greenlee, Mel. 1980. Learning the phonetic cues to the voiced-voiceless distinction: A comparison of child and adult speech perception*. *Journal of Child Language* 7(3). 459–468. <https://doi.org/10.1017/S0305000900002786>.
- Haggard, Mark, Stephen Ambler & Mo Callow. 1970. Pitch as a voicing cue. *The Journal of the Acoustical Society of America* 47. 613–617. <https://doi.org/10.1121/1.1911936>.

- Hogan, John & Anton Rozsypal. 1980. Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. *The Journal of the Acoustical Society of America* 67(5). 1764–1771. <https://doi.org/10.1121/1.384304>.
- House, Arthur. 1961. On vowel duration in English. *The Journal of the Acoustical Society of America* 33(9). 1174–1178. <https://doi.org/10.1121/1.1908941>.
- House, Arthur & Grant Fairbanks. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America*, 25(1). 105–113. <https://doi.org/10.1121/1.1906982>.
- Jaeger, T. Florian. 2008. Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang* 59. 434–446.
- Kessinger, Rachel & Sheila Blumstein. 1998. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics* 26(2). 117–128. <https://doi.org/10.1006/jpho.1997.0069>.
- Klatt, Dennis. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54(4). 1102–1104. <https://doi.org/10.1121/1.1914322>.
- Klatt, Dennis. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5). 1208–1221. <https://doi.org/10.1121/1.380986>.
- Ko, Eon-Suk. 2007. Acquisition of vowel duration in children speaking American English. *Interspeech* 1881–1884.
- Ko, Eon-Suk, Melanie Soderstrom & James Morgan. 2009. Development of perceptual sensitivity to extrinsic vowel duration in infants learning American English. *The Journal of the Acoustical Society of America* 126(5). 134–139. <https://doi.org/10.1121/1.3239465>.
- Krause, Sue. 1982. Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America* 71(4). 990–995. <https://doi.org/10.1121/1.387580>.
- Lisker, Leigh. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33(1). 42–49. <https://doi.org/10.2307/410949>.
- Luce, Paul & John Charles-Luce. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *The Journal of the Acoustical Society of America* 78(6). 1949–1957. <https://doi.org/10.1121/1.392651>.
- Maddieson, Ian & Jack Gandour. 1975. Vowel length before stops of contrasting series. *The Journal of the Acoustical Society of America* 58(S1). S61–S61. <https://doi.org/10.1121/1.2002224>.
- McMurray, Bob, Michael Tanenhaus & Richard Aslin. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86(2). B33–42. [https://doi.org/10.1016/s0010-0277\(02\)00157-9](https://doi.org/10.1016/s0010-0277(02)00157-9).
- Miller, Joanne & Emily Dexter. 1988. Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology. Human Perception and Performance* 14(3). 369–378. <https://doi.org/10.1037//0096-1523.14.3.369>.
- Mugitani, Ryoko, Ferran Pons, Laurel Fais, Christiane Dietrich, Janet Werker & Shigeaki Amano. 2009. Perception of vowel length by Japanese- and English-learning infants. *Developmental psychology* 45(1). 236–247. <https://doi.org/10.1037/a0014043>.
- Naeser, Margaret. 1970. *The American child's acquisition of differential vowel duration. Part 2 of two parts*. Washington: ERIC Clearinghouse.

- Pisoni, David. B., & Jeffrey Tash. 1974. Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics* 15(2). 285–290.
- Port, Robert & Jonathan Dalby. 1982. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics* 32(2). 141–152. <https://doi.org/10.3758/BF03204273>.
- Raphael, Lawrence, Michael Dorman & Donna Geffner. 1980. Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children. *Journal of Phonetics* 8(3). 335–341. [https://doi.org/10.1016/S0095-4470\(19\)31483-4](https://doi.org/10.1016/S0095-4470(19)31483-4).
- Repp, Bruno. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92(1). 81–110. <https://doi.org/10.1037/0033-2909.92.1.81>.
- Sharf, Donald. 1962. Duration of post-stress intervocalic stops and preceding vowels. *Language and Speech* 5(1). 26–30. <https://doi.org/10.1177/002383096200500103>.
- Slis, I. & A. Cohen. 1969. On the complex regulating the voiced-voiceless distinction II. *Language and Speech* 12(3). 137–155. <https://doi.org/10.1177/002383096901200301>.
- Stager, Christine & Janet Werker. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388(6640). 381–382. <https://doi.org/10.1038/41102>.
- Stevens, Kenneth & Dennis Klatt. 1974. Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America* 55(3). 653–659. <https://doi.org/10.1121/1.1914578>.
- Stoel-Gammon, Carol & Eugene Buder. 1999. Vowel length, post-vocalic voicing and VOT in the speech of two-year-olds. *Proceedings of the XIVth International Congress of Phonetic Sciences* 3. 2485–2488.
- Swingley, Daniel & Richard Aslin. 2000. Spoken word recognition and lexical representation in very young children. *Cognition* 76(2). 147–166. [https://doi.org/10.1016/S0010-0277\(00\)00081-0](https://doi.org/10.1016/S0010-0277(00)00081-0).
- Swingley, Daniel & Suzanne Feest. 2019. A cross-linguistic examination of toddlers' interpretation of vowel duration. *Infancy* 24(3). 300–317. <https://doi.org/10.1111/infa.12280>.
- Toscano, Joseph & Bob McMurray. 2012. Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics* 74(6). 1284–1301. <https://doi.org/10.3758/s13414-012-0306-z>.
- Wardrip-Fruin, Carolyn & S. Peach. 1984. Developmental aspects of the perception of acoustic cues in determining the voicing feature of final stop consonants. *Language and Speech* 27(4). 367–379.
- Warren, Paul & William Marslen-Wilson. 1988. Cues to lexical choice: Discriminating place and voice. *Perception & Psychophysics* 43(1). 21–30. <https://doi.org/10.3758/BF03208969>.
- Werker, Janet. 1995. *Exploring developmental changes in cross-language speech perception*. Cambridge, MA: The MIT Press.
- Werker, Janet & Richard Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America* 75(6). 1866–1878. <https://doi.org/10.1121/1.390988>.